Toxicology Summary:

Performance of the Up-and-Down Procedure

1. What is this summary?

This summary is intended to accompany the Up-and-Down Procedure (UDP) test guideline and the AOT425StatPgm program and assist the user in interpreting its output. The conclusions in this summary are derived from applying the test procedure to many repetitions of test simulations. More extensive treatment of the simulations can be found in the report "Simulation Results for the AOT425StatPgm Program", which can be downloaded from the USEPA website.

The UDP guideline uses the maximum likelihood estimation method for calculation of the LD50 and profile likelihood (PL) methods for estimation of most confidence intervals. A wide confidence interval indicates that there is more uncertainty associated with the estimated LD50. In this case, the reliability of the estimated LD50 is low and the usefulness of the estimated LD50 may be marginal. A narrow interval indicates that there is relatively little uncertainty associated with the estimated LD50. In this case, the reliability of the estimated LD50 is high and the usefulness of the estimated LD50 is good.

This summary addresses choice of dose progression and initial dose level for the UDP and describes the performance of the test under a variety of circumstances. The statistical methods applied will depend upon the case into which the test response patterns fall.

2. Adjusting the Dose Progression and Initial Dose.

All available information on the test substance should be considered by the testing laboratory prior to conducting the study in order to select the best initial dose and dose progression or spacing. Ideally, initial dose should be just below the prior estimate of the LD50 of the test material and dose progression should be based on the toxicologist's best estimate of slope of the dose-response curve, or *sigma*, the standard deviation of the log normal curve of animal responses to the test chemical. (Slope = 1/*sigma*).

In the absence of such information, the UDP guideline recommends a default starting dose of 175 mg/kg and use of half log units corresponding to a dose progression of 3.2. These default settings will accommodate a variety of situations, including chemicals with slopes as low as 2.0. However, most industrial chemicals tend to have steeper slopes, up to a value of 8 or higher. Many pesticides and chemicals whose toxicity is receptor mediated may exhibit slopes as low as 2 or 2.5.

Flexible stopping rules governing the number of animals tested allow the UDP to be applied to test materials with a wide range of slopes. However, for best performance of the UDP, the dose progression used should be based on an accurate estimate of *sigma*. The following two cases describe the outcome when an accurate estimate of *sigma* is not available. In addition, to account conservatively for any bias in the LD50 estimate, it is essential that dosing be initiated below the actual LD50.

(I) Assumed *sigma* << true *sigma*:

When the assumed *sigma* (i.e., the *sigma* on which the dose progression is based) is much smaller than the true *sigma* of the actual test population, the estimated LD50 may be "biased" in the direction of starting dose. For example, if the starting dose is less than the true LD50 of the test population, the estimated LD50 will generally be below the true LD50. Also, if the starting dose is greater than the true LD50 of the test population, the estimated LD50 will tend to be greater than the true LD50. To

minimize the chance of overestimating the LD50 due to this bias, the UDP guideline recommends a choice of starting dose just below the assumed LD50.

(ii) Assumed *sigma* >> true *sigma*:

If the assumed *sigma* on which the dose progression is based is much larger than the true *sigma* of the test population, the LD50 can be estimated only within a range. (This is Case 3 described below.)

3. Response Patterns.

Data gathered under the UDP fall into one of five animal response patterns. The five types of animal response patterns, referred to as Case 1 - Case 5 below, can be distinguished for the purpose of describing the performance of the UDP. (These cases are also summarized in Table 1.) These cases can be distinguished by looking at the experimental outcome (survival or death) as reflected in the AOT425StatPgm final Data Grid or final Report windows. In considering these cases, note that doses can be repeated more than once in the course of sequential dosing.

Case 1. No Dose-response:

Case 1 has three possible response patterns: (a) all animals responded at all doses tested, (b) no animals responded at any dose tested, or (c) the geometric mean dose is lower for animals that respond than for animals that did not respond.

Case 1(a) indicates that the LD50 is likely to be lower than the lowest dose tested. In this case, the lowest dose tested can be considered an upper boundary to the actual value for the median toxic level or LD50. Therefore, the LD50 is likely to be below any dose tested, and in particular below the lowest dose tested. The AOT425 software does not provide a confidence interval for this case.

Case 1(b) indicates that the LD50 is likely to be greater than the highest dose tested. In this case, the highest dose tested is essentially a limit dose and the LD50 is likely to be above that dose. The AOT425 software does not provide a confidence interval for this case.

Case 1(c) is extremely unlikely to occur since the negative dose-response curve implied by this response pattern is not characteristic of toxic chemicals. The AOT software does not provide a confidence interval for this case.

Case 2:

Case 2 is when the greatest dose with some nonresponses (at least some survivors) is greater than the lowest dose with at least one response (at least one animal dies). That is, the case shows partial responses at certain doses (i.e. some but not all animals respond at those doses). Typically, partial responses will occur at more than one dose level. For case 2, maximum likelihood estimates of the LD50 and finite 95% confidence intervals use the profile likelihood method..

Case 3:

Case 3 has no partial responses. All animals die at higher doses and at lower doses, all animals survive. This implies that the LD50 is between the highest dose with no response and the lowest dose where complete responses occur. This case occurs most often when the dose spacing is large relative to the actual variance of the lethality normal curve. In this case, any value between these two doses might be the true LD50. Although the AOT425StatPgm provides an estimate of LD50 by taking the median value between these doses, the test response should be regarded as providing a range estimate of lethality, and hence the AOT425 point estimate is artificial. Although a confidence interval *per se* is not

computed, simulations suggest that, in effect, the two doses can occur as a 95% confidence interval. Therefore, this range is provided by the AOT425 software.

Case 4:

Case 4 has a single dose showing partial response (i.e. some, but not all animals die at a single dose). The partial response dose is between the doses associated with 0% response and those associated with 100% response.

For Case 4, the LD50 estimate provided by the AOT425 software is the single dose at which the partial response occurred. Confidence intervals are calculated using the profile likelihood method.

Case 5:

Case 5 includes two possible situations.

One possibility shows partial response at the highest dose and no responses at lower doses. This suggests that the LD50 is around the highest dose tested or possibly higher.

The second situation shows partial response at the lowest tested dose and complete responses at higher doses. Here, the LD50 is likely to be at or below the lowest dose tested.

For Case 5, the LD50 estimate of the software will be the dose with the partial response. The confidence interval will be calculated using the profile likelihood method.

4. How is the LD50 Estimate Calculated?

The LD50 estimate is calculated using the maximum likelihood method unless the response pattern falls into an exceptional case (included in some of the response patterns described above). All deaths, whether immediate or delayed or humane kills, are incorporated for the purpose of the maximum likelihood analysis.

In performing the maximum likelihood calculation, an estimate of *sigma* of 0.5 (corresponding to the default dose progression factor) is used unless a better generic or case-specific value is available and has been specified by the user. If a better value of *sigma* is available, the dose spacing should be adjusted accordingly, as instructed in the guideline.

Under some circumstances, statistical computation will not be possible or will likely give erroneous results. Special means to determine/report an estimated LD50 are available for these circumstances as follows:

(a) If testing stopped because a boundary dose was tested repeatedly, or if the upper bound dose ended testing, then the LD50 is reported to be above the upper bound.

(b) If all the dead animals have higher doses than all the live animals (or if all live animals have higher doses than all the dead animals, although this is practically unlikely), then one may have reasonable confidence that the LD50 is between the doses for the live and the dead animals. These observations give no further information on the exact value of the LD50. Still, a maximum likelihood LD50 estimate can be made provided there is a value for *sigma*. When the actual value of *sigma* is not available, the AOT425 software calculates a maximum likelihood estimate based on the dose spacing used. However, when this type of response pattern is seen, it is clear that the actual *sigma* is much smaller than the value used to set the dose progression. Therefore, point estimate LD50 values are artificial. Rather, the yield of the

test is a *range* for lethality. If a closely related substance is tested, testing should proceed with a smaller dose progression.

(c) If the live and dead animals have only one dose in common and all the other dead animals have higher doses and all the other live animals lower doses, or vice versa, then the LD50 estimate equals their common dose. If a closely related substance is tested, testing should proceed with a smaller dose progression.

If none of the above situations occurs, then the LD50 estimate is calculated using the maximum likelihood method.

5. How is the Confidence Interval Calculated?

Following the main test and estimated LD50 calculation, it may be possible to compute a confidence interval for the LD50. Any of these confidence intervals provides valuable information on the reliability and utility of the main test that was conducted. A confidence interval can be viewed as providing plausible bounds on the value of the LD50 based on the data collected in the particular study. Whereas, point estimation results in a single value estimate for the LD50, confidence interval estimation is expressed in a lower and upper bound for an interval that has a known probability or confidence of containing the true value of the LD50.

The UDP guideline calls for use of profile likelihood methods to calculate confidence intervals. Calculation of the profile likelihood requires maximizing the likelihood (function) while holding the term for the LD50 at a fixed assumed value. At each fixed assumed LD50, the likelihood will be maximized by some particular value of the slope. Calculation of the profile likelihood confidence intervals requires calculating the profile likelihood for different values of fixed assumed LD50s with their corresponding profile maximizing slopes and finding the value for which the profile likelihood equals a critical value. This is a computationally-intensive procedure. Consequently, these are incorporated into the AOT425 software.

A measure of the performance of a confidence interval procedure is the coverage. The coverage is the probability that a calculated confidence interval, based on a sample, encloses the true LD50 for a sample population. Simulations suggest that coverage falls below 95% when the slope is shallow and above 95% when slopes are very steep. The UDP statistical algorithm is designed to compute 90%, 95%, or 99% profile likelihood confidence intervals. However, the algorithm is not exact but approximate, so that in some situations, the confidence interval will not provide the desired coverage or may provide more than the desired coverage.

Results of the simulations can be summarized using the following rules of thumb. It appears that the nominal 95% confidence interval will have coverage at least 90% if the slope is 2-4 or greater (*sigma* 0.25-0.5 or smaller). (For most situations, the coverage will be better than 90% if the slope is 2 or greater.) Coverage will be 80% or better if the slope is at least 1. For slopes as low as 0.5 (the lowest slope assumed in simulations) the coverage may be as low as 70%. It is suggested that, in such a situation, it will be clear to the experimenter that narrower bounds can be obtained by using doses that are more closely spaced, and that probably no type of confidence interval would be narrower than the current dose spacing.

Depending on the outcome of the test, one of three different types of confidence intervals for the true LD50 is calculated:

(a) <u>When the UDP provides a point estimate of the LD50:</u> When at least three different doses have been tested and the middle dose has at least one animal that survived and one animal that

died, a profile-likelihood-based computational procedure is used to obtain a confidence interval that is expected to contain the true LD50 95% of the time. However, because small numbers of animals are expected to be used, the actual level of confidence is generally not exact (Jennison and Turnbull, 2000). The random stopping rule in the UDP improves the ability of the test overall to respond to varying underlying conditions, but also causes the reported level of confidence and the actual level of confidence to differ somewhat (Shiryaev and Spokoiny, 2000).

(b) When the UDP provides a range estimate of the LD50: If all animals survive at or below a given dose level and all animals die when dosed at the next higher dose level, a confidence interval is calculated that has as its lower limit the highest dose tested where all the animals survive and has as its upper limit the dose level where all the animals died. This interval is labeled as "approximate." The exact confidence level associated with this interval cannot be specifically determined from the data obtained in the test. However, because this type of response would ordinarily occur when the dose-response is steep, in most cases, the true LD50 is expected to be contained within the calculated interval or will be very close to it.

(c) When the dose-response curve is flat or the standard deviation is large: In some instances, confidence intervals are reported as infinite, through including either zero at the lower end or infinity at the upper end, or both. Such intervals may occur, for example, when the slope of the dose-response is relatively flat or relatively uncertain.

When all doses show no response (i.e., animals survive at every dose tested) the highest dose tested is equivalent to a limit dose above which the LD50 is expected to fall. In this case, the conventional concept of confidence interval is not applicable and the AOT software does not provide a confidence interval.

References.

Jennison, C. and B.W. Turnbull, 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC; Boca Raton, FL.

Shiryaev, A.N. and V.G. Spokoiny, 2000. Statistical Experiments and Decisions. Statistical inference for autoregressive models of the first order asymptotic theory. Vol. 8, Chapter 5. World Scientific Publ., London, Singapore.

Table 1. Outcomes of the Up-and-Down Procedure: Cases and Confidence Intervals.

Case #	Definition of Case	Approach Proposed	Possible Findings
1	**No positive dose-response association.** 1a) all animals tested in the study responded, or 1b) none responded, or 1c) the geometric mean dose is lower for animals that responded than for animals that did not respond.	LD50 cannot be calculated. Confidence interval not applicable.	Possible inferences: 1a) LD50 < lowest dose; 1b) LD50 > highest dose; 1c) reverse dose-response curve; unlikely test outcome. In case 1b, the highest dose tested is equivalent to a limit dose.
2	**Multiple partial responses.** One or more animals responded at a dose below some other dose where one or more did not respond. The conditions defining Case 1 do not hold. (The definition of Case 2 holds if there are 2 doses with partial responses, but holds in some other cases as well.)	Maximum likelihood estimate and profile likelihood computations of confidence interval are straightforward.	The LD50 can be estimated and its confidence interval calculated.
3	**No intermediate response fractions.** One or more test doses is associated with 0% response and one or more is associated with 100% response (all of the latter being greater than all of the former), and no test doses are associated with a partial response.	Lower bound = highest test dose with 0% response. Upper bound = lowest test dose with 100% response.	High confidence that the true LD50 falls between the two bounding doses. Any value of LD50 between highest dose with 0% response and lowest dose with 100% response is equally plausible.
4	**One partial response fraction, first subcase.** An intermediate partial response is observed at a single test dose. That dose is greater than doses associated with 0% response and lower than doses associated with 100% response.	The LD50 is set at the single dose showing partial response and its confidence interval is calculated using profile likelihood method.	The LD50 can be estimated and its confidence interval calculated.

5	**One partial response fraction, second subcase.** There is a single dose associated with partial response, which is either the highest test dose (with no responses at all other test doses) or the lowest test dose (with 100% response at all other test doses).	The LD50 is set at the dose with the partial response. A profile likelihood confidence interval is calculated and may be finite or infinite.	The true LD50 could be at the boundary of the testing range with more or less confidence.